Free Choice Questions (Draft)

Tomasz Klochowicz*

January 24, 2024

Abstract

Polar questions like "May I go to the park or to the beach?" give rise to inferences similar to Free Choice Permission. The *Yes* answer to these questions corresponds to the permission to freely choose between going to the park and going to the beach. *No* corresponds to Dual Prohibition, i.e., prohibition to go to either place. We empirically tested these intuitions. We will indicate how the collected data can allow us to establish the source of these inferences and compare the findings to predictions made by current theories of Free Choice extended with question semantics. The collected data poses a challenge to the semantic and scalar approaches to free choice and supports non-scalar pragmatics as a uniform solution to the *free choice* puzzle.

1 Introduction

Georg von Wright (1968) and Hans Kamp (1973) observed that speakers accept inferences like (1) contrarily to the predictions of classical logic.[1] They called these sentences *Free Choice Permission* because asserting them implies that the interlocutor can freely choose between the two proposed options. The aim of this paper is to determine whether various theories explaining this *free choice* inference generalise to its inquisitive version (2), which we will call a *Free Choice Questions* (*FCQ*). The theoretical part of this paper focuses on exploring the possible ways of modelling *FCQ*s. The empirical part reports on an experiment which tests the predictions of those theories regarding the meaning of the response particles: *Yes* and *No* as answers to *FCQ*s. Intuitively, the response particles behave as represented in (2). In Section 3, we provide a detailed description of the empirical confirmation of this intuition.

(1) You may keep a dog or a cat in this apartment. $\Diamond(\alpha \vee \beta)$
$\rightsquigarrow$ You may keep a dog and you may keep a cat (but maybe not both). $\Diamond\alpha \wedge \Diamond\beta$

(2) A: May I keep a dog or a cat in this apartment? $?\Diamond(\alpha \vee \beta)$
B: Yes $\overset{?}{\rightsquigarrow}$ You may keep a dog and you may keep a cat. $\Diamond\alpha \wedge \Diamond\beta$
B: No $\overset{?}{\rightsquigarrow}$ You may not keep a dog and you may not keep a cat. $\neg\Diamond\alpha \wedge \neg\Diamond\beta$

Why is the inference in (1) problematic for classical logic? If the following *FC* principle seems to hold in natural language reasoning, can't we just add it as an axiom to our logical system?

FC: $\Diamond(\alpha \vee \beta) \rightarrow \Diamond\alpha$

*I would like to thank my supervisors Maria Aloni and Floris Roelofsen for helpful feedback and comments throughout the work on this material. I am also very grateful to Søren Brinck Knudstorp, Marco Degano, Dean McHugh, Sonia Ramotowska, Tom Roberts, Jacopo Romoli, Jakub Szymanik, Yasu Sudo, audiences of the NihiL Seminar, the LOT summer school and PALLMYR-XIII for discussion and comments on this work. All remaining mistakes are my own. This work is supported by Nothing is Logical (NihiL),an NWO OC project (grant no 406.21.CTW.023)

[1] For the empirical confirmation of this fact see e.g. Chemla and Bott (2014), Cremers et al. (2017) or Marty et al. (2021)

This simple extension is not possible since classical (deontic) logic, allows for disjunction introduction under deontic modality: by monotonicity of the deontic modal operator ($\Diamond$): $\Diamond\alpha$ implies, $\Diamond(\alpha \vee \beta)$. Adding *FC* as an axiom to classical logic would validate the reasoning below, which states that if something is allowed, then anything is allowed:

You may keep a dog.	$\Diamond\alpha$	(*Assumption*)
You may keep a dog or a crocodile.	$\Diamond(\alpha \vee \beta)$	(*Modal Addition*)
$\rightsquigarrow$ You may keep a crocodile.	$\Diamond\beta$	(*FC*)

The issue has practical importance since, as noticed by Aher (2013), many legal documents contain statements in the form of *free choice*. We can find them in the Universal Declaration of Human Rights (Article 8), the Constitutions of the USA (e.g. Article 1 Section 4 and 7), the Netherlands (e.g. Articles 82 and 134) and Poland (e.g. Article 42.2) Since the classical interpretation of logical connectives is used in legal reasoning, accepting the *FC* principle would make these legal documents vacuous, and not accepting it leads to misinterpretation of the established laws. Consider the following passage:

(3) *If you pass [the driving license test] you may ride a motorcycle up to 125 cc with power output up to 11 kW, or a motor tricycle with power not exceeding 15 kW.*[2]

If we accept *FC* in its interpretation, without blocking modal addition, we can infer that with the driving license, we may drive any vehicle e.g. a tank. If we do not accept *FC* we cannot be sure if we are allowed to drive both motorcycles and motor tricycles.

Aher (2013) mentioned a simple, legal solution proposed by the New York Court of Appeals: *"Generally, the words "or" and "and" in a statute may be construed as interchangeable when necessary to effectuate legislative intent"*[3]. However, as pointed out by Aher (2013), deriving legislative intent should be avoided if the same reasoning can be made using the literal meaning of the law. So maybe we can define the conditions where they can be interchanged and limit them to occurrences under a deontic modality? This solution will not work either:

To block *modal addition* we could, somewhat naively, claim that under a deontic modality, disjunction behaves like a conjunction, and we cannot freely add another disjunct. However, this move leads to issues with the Dual Prohibition (*DP*) inference (Alonso-Ovalle, 2006; Aloni, 2022). Utterances like (4) behave classically, i.e. prohibition of disjunction implies the prohibition of each disjunct.[4] Our *ad hoc* solution would not work here, as it would predict that (4) is a negation of a *free choice* sentence, which would only imply that one of the two disjuncts is not permitted. Alonso-Ovalle (2006) and Aloni (2022) indicated that this imbalance between *FC* and *DP* will cause problems to all purely semantic approaches to *free choice*.

DP: $\neg\Diamond(\alpha \vee \beta) \rightarrow \neg\Diamond\alpha$

(4) You may *not* keep a dog or a cat in this apartment. $\neg\Diamond(\alpha \vee \beta)$

$\rightsquigarrow$ You may *not* keep a dog and you may *not* keep a cat. $\neg\Diamond\alpha \wedge \neg\Diamond\beta$

Since the documents mentioned above contain laws formulated using Dual Prohibition (e.g. (5)), the *ad hoc* solution would still lead to possible misinterpretation. Can we extend it by adding that *and* and *or* can be interchanged only if the deontic modality does not occur under negation? This will cause problems with legal questions in the form of a Free Choice Question like (6) and (7). These questions are used to delimit the scope of a case or as prejudicial questions. Especially in case law, their interpretation may be crucial to a case. In particular, it should be clear what is permitted and what is prohibited once such a question is answered.

(5) *This right* may not *be invoked in the case of prosecutions genuinely arising from non-political*

[2] The highway code of the UK p.51 `https://www.highwaycodeuk.co.uk/download-pdf.html`, access: 30.05.2023

[3] 425 U.S. at 410 n. 11

[4] See e.g. Marty et al. (2021) for empirical confirmation of this inference.

crimes or *from acts contrary to the purposes and principles of the United Nations.*[5]

(6) ***The scope of the case will be at once made manifest by the two questions*** *which were certified for solution. First:* May a patentee *or his assignee license another to manufacture and sell a patented machine and by a mere notice attached to it* limit its [patent's] use by the purchaser or by the purchaser's lessee*, to films which are no part of the patented machine, and which are not patented?* [...][6]

(7) *May a State prohibit children or foreigners from circulating petitions [...]?*[7]

Observe that *FCQ*s are not specific to legal discourse. They can be used in everyday conversations to ask for permission (8), but also to ask to report what is permitted (9).

(8) May I have ice cream or cake?

(9) Am I allowed to visit the Rijks or Nemo with my Museumkaart?

In this paper, we focus on the simplest answers to *FCQ*s, namely on response particles: *Yes* and *No*. We can say that such a question divides the logical space into (at least) two parts. The positive one is picked out by the *Yes* particle and the negative is picked out by *No*. Possible meanings of the response particles, as answers to an *FCQ*, are represented in Figure 1.

Figure 1(a) represents the case where *Yes* corresponds to the literal meaning of the *free choice* statement (at least one of the two choices is allowed), and *No* to its complement, i.e. to *Dual Prohibition* (Neither is allowed). Figure 1(b) represents the case where *Yes* corresponds to *free choice* (Both allowed) and*No* to its complement (At least one not allowed). The last figure (1(c)) corresponds to the case where *Yes* corresponds to *FC* and *No* to *DP*, leaving the parts in which only one of the choices is allowed outside the reach of the response particles.

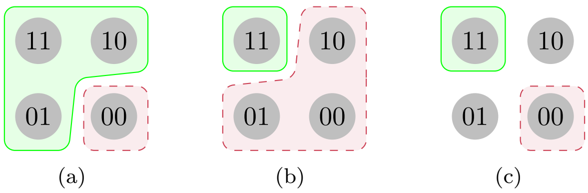


Figure 1: Possible interpretations of response particles. The labels indicate which disjuncts are allowed (1) and not allowed (0). Solid lines correspond to the "Yes" answer and dashed lines to the "No" answer. For instance in question (2): "May I keep a dog or a cat?", in the state in the top left corner (11) keeping a dog is allowed and keeping a cat is allowed; in the bottom right state (00) keeping neither dog nor cat is allowed.

In the next section, I present a typology of solutions to the *free choice* puzzle and indicate how they solve the issues of *FC* and *DP*. Moreover, I will discuss the predictions of these theories regarding *FCQ*s. In Section 3, I will report on our experiment regarding the meaning and processing of the response particles as answers to *FCQ*s. Section 4 will compare the predictions of the existing theories of *FC* with the results of the experiment.

2 Theories of Free Choice Questions

The problem of declarative *free choice* received various analyses over the years. The aim of this section is to discuss the predictions of existing theories regarding the inquisitive version: the *FCQ*s. I will present a typology of solutions to the *free choice* puzzle and indicate how they can be extended with a theory of questions to model *FCQ*s. From these extended theories I will derive the predictions regarding the meaning of the response particles as answers to *FCQ*s.

[5]Universal Declaration of Human Rights, Article 14; `https://www.un.org/en/about-us/universal-declaration-of-human-rights` access: 31.05.2023

[6]Boston Store of Chicago v. American Graphophone Co, 246 U.S. 8 (1918)

[7]Buckley v. American Constitutional Law, 525 U.S. 182 (1999)

2.1 Theories of Free Choice

Following Aloni (2022) we can group the solutions to the *free choice* puzzle into three categories:

Semantic theories redefine semantic element(s) of classical logic involved in the *free choice* inferences ('$\vee$', '$\Diamond$' or '$\neg$'), and explain *FC* in terms of the non-classical behaviour of these element(s). Semantic theories were proposed by, among others, Simons (2005), Aloni (2007, 2016, 2018) and Barker (2010). As we indicated above, these solutions may have trouble accommodating *FC* and *DP* at the same time because of the asymmetrical behaviour of those inferences. As a representative, we will discuss Deontic Inquisitive Logic proposed by Nygren (2022).

Scalar theories keep the logic intact but postulate a scalar inference, which allows to derive *FC* and *DP* without paradoxical consequences. Note that these theories may explain *DP* "for free", as it is a valid inference in classical logic. The proposed pragmatic mechanism should explain why we take *free choice* inferences to be (conversationally) valid without invalidating *DP*. As an example of a scalar theory, we will discuss the Exhaustification approach proposed by Fox (2007) and refined by Bar-Lev and Fox (2020). These theories were designed as a solution to issues of the Neo-Gricean theories proposed by Gazdar (1979) or Kratzer and Shimoyama (2002) involving the usage of Grice's maxims.

Non-scalar pragmatics postulate the existence of a pragmatic principle, explaining *free choice* using either classical or non-classical semantics. These general tendencies in human reasoning are independently motivated. As two representatives of these approaches, we will discuss the homogeneity approach in the spirit of Goldstein (2019) and the neglect-zero approach proposed by Aloni (2022).

In the remainder of this section, we will describe theories representative of these strategies and discuss their solutions in more detail. Moreover, we will indicate what are the predictions of these theories regarding the meaning of response particles as answers to *FCQ*s. However, before doing so, I need to focus a bit on the way they will be extended to allow them to model questions.

2.2 Theories of Questions

There are many proposals which aim to thoroughly capture the meaning and behaviour of questions and their relation to declaratives (answers). The most prominent theories in the literature follow one of the two ideas about what questions are: Hamblin (1976) and Karttunen (1977) see questions as sets of propositions, and Groenendijk and Stokhof (1984) as partitions of the logical space.

Approaches of Dayal (1996) and Fox (2018, 2020) aim to combine the two, by designing a mechanism which allows to determine the partition induced by a set of classical propositions. The core idea is to derive the pragmatically relevant partitions from semantically unproblematic propositions. On the other hand, Ciardelli et al. (2018) propose to interpret both declarative propositions and questions as sets of sets of possible worlds (instead of just sets of worlds), which allows for a uniform treatment of declaratives and questions.

I will not argue for or against any of these theories as the best choice to capture *FCQs*. In fact, the choice of a theory of questions should not matter too much. The predictions of a theory regarding the meaning of response particles should be fully determined by the way it treats declaratives.

Moreover, since the scope of this paper is limited to polar questions, we can take any theory which has the ability to take a declarative P and return a polar question: "Is P true?". Since theories of questions differ mainly in empirical predictions regarding constituent questions, this should not be problematic for any theory. We will model this behaviour by using a question operator: "?".

The operator applied to a formula φ should return a positive part corresponding to *Yes* and a negative part corresponding to *No*. The operator should do so in a systematic way so that it is possible for any given declarative to logically deduce the parts returned by ?. It is sensible to assume that these parts are the declarative and its negation. We will use the logical form of questions from Ciardelli et al. (2018) extended with the highlighting mechanism by Roelofsen and

Farkas (2015), as it allows for uniform representation of questions and declaratives, but any other theory of questions, which has an operator satisfying the aforementioned criteria would be fine:

$$?\varphi \equiv \underbrace{\varphi}_{\text{Yes}} \vee\!\!\!\vee \underbrace{\neg\varphi}_{\text{No}}$$

Consider the logical form of a free choice question (*May I keep a dog or a cat?*): $?\Diamond(\alpha \vee \beta)$. All the accounts of declarative *free choice* have already defined the interpretation of all the connectives (i.e. "$\Diamond$" and "$\vee$") and the mechanisms involved in the processing of the part under the question operator. Let FC denote the logical form of *free choice* postulated by a theory, then the general logical form of a Free Choice Question can be analysed as follows:

$$?\text{FC} \equiv \underbrace{\text{FC}}_{\text{Yes}} \vee\!\!\!\vee \underbrace{\neg\text{FC}}_{\text{No}}$$

Observe that the disjuncts of this formulation correspond to *free choice* and *dual prohibition.* We can see that if a theory of *free choice* makes predictions about *FC* and *DP*, it should also have predictions for FCQs.

Below, I will discuss possible extensions of existing theories of Free Choice, which allow them to capture polar questions. We will indicate which predictions they make regarding the meaning of response particles as responses to Free Choice Questions. Recall the possible meanings represented in Figure 1. For each theory, we will indicate which pattern they predict. Moreover, we will make some indications regarding the processing characteristics that they predict.[8]

2.3 Semantic theories

Semantic theories do not commit to any pragmatic factors, but try to solve the issue of *free choice* by postulating non-classical semantics for logical expressions.

2.3.1 Inquisitive logic with a classical modal operator

In the introduction, we have seen that classical logic is not fit to model declarative *free choice.* However, it may still give correct predictions regarding the inquisitive version. Let's consider an extension of classical logic by Ciardelli et al. (2018), which allows to model questions. Let the formula "$\varphi \vee\!\!\!\vee \psi$" express a question: *whether φ or ψ*. It is satisfied when at least one of the disjuncts is satisfied:

$M, s \models \varphi \vee\!\!\!\vee \psi$ iff $M, s \models \varphi$ or $M, s \models \psi$ (Inquisitive disjunction)

Ciardelli (2016) proposes the standard interpretation of an existential modal, which is defined as follows:

$M, s \models \Diamond\varphi$ iff for all $w \in s : [\varphi] \cap R[w] \neq \emptyset$ (Simple modality)

To use the general form of an *FCQ* we also need negation. In Inquisitive Logic, $\neg\varphi$ is supported by s if φ is not supported in s.

$M, s \models \neg\varphi$ iff for all $w \in s$ $M, w \not\models \varphi$

Where $[\varphi]$ is the set of worlds which support φ (the *truth set*) and $R[w]$ is the set of worlds *seen* by w through the relation R. We can now use our general form of an *FCQ* from the previous section to represent it in inquisitive logic:

$$?\Diamond(\alpha \vee\!\!\!\vee \beta) \equiv [\underbrace{\Diamond(\alpha \vee\!\!\!\vee \beta)}_{\text{Yes}} \vee\!\!\!\vee \underbrace{\neg\Diamond(\alpha \vee\!\!\!\vee \beta)}_{\text{No}}]$$

Nygren (2022) observes that this definition of $\Diamond$ implies that under the modality, inquisitive disjunction behaves exactly like classical disjunction since it is evaluated with respect to possible

[8] In the next section, we will only indicate the tendencies regarding processing that are predicted by the theories. Concrete predictions regarding the results of the experiment will be discussed in Section 3.

worlds (and not states). Therefore, the parts of *FCQ* are behaving classically. The positive part corresponds to a classical proposition $\Diamond(\alpha \vee \beta)$, which states that at least one of the choices is allowed, and the negative part is its classical negation, which states that neither of the choices is allowed.

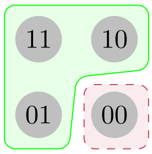


Figure 2: Response pattern predicted by inquisitive logic with a classical modal operator. The labels indicate which disjuncts are allowed (1) and not allowed (0). Solid lines correspond to the "Yes" answer and dashed lines to the "No" answer.

2.3.2 Deontic Inquisitive Logic

As an example of a semantic theory of *free choice*, I will discuss Deontic Inquisitive Logic by Nygren (2022), who uses non-classical definitions of disjunction and existential modality, which allows for satisfaction of *FC*. This strategy is typical for a semantic account.

Deontic Inquisitive Logic is state-based, as it evaluates the formulas with respect to information states (sets of possible worlds) and not individual possible worlds. Nygren (2022) proposes to take disjunction to be inquisitive in the spirit of Ciardelli et al. (2018) (see above).

Nygren (2022), following the suggestions by Ciardelli (2016) proposes a new definition for $\Diamond$. Formula $\Diamond\varphi$ expresses that for any world w in the evaluation state, s each *alternative* of φ is true in some world w' accessible from w.[9] The formula $\varphi \mathbin{\vee\!\!\!\vee} \psi$ has two alternatives, namely the φ and ψ: $Alt_M(|\varphi \mathbin{\vee\!\!\!\vee} \psi|_M) = \{|\varphi|_M, |\psi|_M\}$.[10]

$M, s \models \Diamond\varphi$ iff for all $w \in s$, for all $Y \in ALT_M(\varphi) : Y \cap R[w] \neq \emptyset$

Recall the formulations of *FC* and *DP* from page 1, where α and β correspond to the two choices. According to the refined definition of modality, $\Diamond(\alpha \mathbin{\vee\!\!\!\vee} \beta)$ is supported at s if for any $w \in s$ α is true in some world accessible from w and β is true in some world accessible from w. From there, it is easy to conclude that $\Diamond\alpha$ is supported by s and thus *FC* holds in DIL:

Proof. (*FC*) If $M, s \models \Diamond(\alpha \mathbin{\vee\!\!\!\vee} \beta)$ then for all $w \in s$, for all $Y \in \alpha \mathbin{\vee\!\!\!\vee} \beta : Y \cap R[w] \neq \emptyset$. Since $\alpha \in Alt_M(\alpha \mathbin{\vee\!\!\!\vee} \beta)$ then for all $w \in s$: $\alpha \cap R[w] \neq \emptyset$. Thus $M, s \models \Diamond\alpha$. □

Negation in this framework is defined exactly the same as in Inquisitive Logic (see above). Observe that there are several ways in which the premise of *FC*: $\Diamond(\alpha \mathbin{\vee\!\!\!\vee} \beta)$ can lack support in s. In particular, this can happen when there is a $w \in s$ such that exactly *one* of the alternatives is false in any world accessible from w, while the other is still true. This treatment of negation yields failure of *Dual Prohibition*:

Proof. (*DP* Failure) Suppose (w.l.o.g) that $M, s \models \Diamond\alpha$ and that $M, s \models \neg\Diamond\beta$, where α and β are not inquisitive. Then there is a world $w \in s$ such that $\beta \cap R[w] = \emptyset$. Since $\beta \in ALT(\alpha \mathbin{\vee\!\!\!\vee} \beta)$ then there is a $w \in s$ and $Y \in ALT_M(\alpha \mathbin{\vee\!\!\!\vee} \beta) : Y \cap [w] = \emptyset$. Thus $M, s \not\models \Diamond(\alpha \mathbin{\vee\!\!\!\vee} \beta)$, so $M, s \models \neg\Diamond(\alpha \mathbin{\vee\!\!\!\vee} \beta)$ □

The logical form of an *FCQ* is exactly the same as in inquisitive logic (see above), but it has different truth conditions because of the difference in the definitions of $\Diamond$. The positive part corresponds to *free choice*, i.e. the case where both choices are allowed, and the negative part to its (classical) negation, i.e., the case where at least one choice is not allowed.

[9] Note that this is not the notion of alternatives used for scalar implicatures, but alternatives in the sense of Ciardelli et al. (2018). Please consult Nygren (2022)'s paper for more details.

[10] For brevity of this explanation I omit here the issue with subordinate alternatives (such that one is a proper subset of the other), as we will not deal with Hurford disjunctions in this paper. However, in further research, it may be interesting to analyse their assertability in questions.

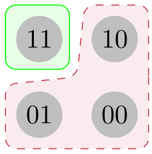


Figure 3: Response patterns predicted by deontic Inquisitive Logic and other purely semantic theories of *free choice*.

Every semantic theory of *free choice* relies on an assumption that all the relevant inferences follow from the literal meaning of asserted statements. Therefore they predict that *FC* and *DP* should be not more difficult than other literal meaning inferences – that there should be no difference in processing. Therefore, answering Free Choice Questions and analysing conversations involving them should be similar to processing any other polar question with similar syntactic complexity.

2.4 Scalar theories

Alonso-Ovalle (2006) hinted that the asymmetry between the behaviour of disjunction in *FC* and *DP* suggests a pragmatic source of *free choice*, as such behaviour is characteristic of scalar implicatures, which are not computed in negative environments as in (10).

(10) a. No one is allowed to keep a dog or a cat here.
$\rightsquigarrow$ No one may keep a dog, and no one may keep a cat.
b. I doubt that you are allowed to keep a dog or a cat here.
$\rightsquigarrow$ I doubt that you may keep a dog and I doubt that you may keep a cat.

Grammatical theories propose to model scalar reasoning using a covert exhaustivity operator: $Exh()$. The operator exhausts the set of alternatives, which is computed syntactically in the style of Sauerland (2004) or in terms of complexity as described by Katzir (2007).[11] Bar-Lev and Fox (2020) following Fox (2007) propose to use the following exhaustification algorithm:

1. Take the prejacent and compute the set of alternatives
2. Take all maximal sets of alternatives, that can be assigned *false* with the prejacent:
3. *Innocent Exclusion:* Exclude the intersection of those sets
4. Take all maximal sets of alternatives, that can be assigned *true* with the prejacent and negations of excluded alternatives.
5. *Innocent Inclusion:* Include the intersection of those sets.

To see how this algorithm solves the issues of *FC* and *DP* the reader can consult the paper by Bar-Lev and Fox (2020). The core idea is that in the positive case $\Diamond\alpha$ and $\Diamond\beta$ can be innocently included, while in the negative case $\neg\Diamond(\alpha \vee \beta)$ is already the strongest expression in its set of alternatives.

Since the grammatical view by Bar-Lev and Fox (2020) can accommodate both *FC* and *DP*, it would seem that it should easily predict that they constitute the positive and negative part of an *FCQ*, respectively. However, a theorist who wants to model *FCQ*s using exhaustification needs to choose whether they take the constituent under the question operator to be exhaustified or not. If it is, then the positive part corresponds to *free choice* (both allowed), but the negative part to its classical negation. If the declarative is not exhaustified then the approach has the same predictions as classical logic:[12]

[11] Note that the notion of *alterntive* is used here in a different way than in section 2.3.1

[12] Observe that the theory of question by Fox (2020) runs into the same issue. The prejaceant of cell identification must be a singleton since the resulting partition needs to be binary (follow the yes/no pattern of a polar question). Again the only reasonable choices are the exhaustified and bare versions of the declarative $\Diamond(\alpha \vee \beta)$. Applying the formalism from that paper yields the same predictions as indicated above. In the formulation inspired by inquisitive

$$?Exh(\Diamond(\alpha \vee \beta)) \equiv \underbrace{Exh(\Diamond(\alpha \vee \beta))}_{\text{Yes}} \mathbin{\vee\!\!\!\vee} \underbrace{\neg Exh(\Diamond(\alpha \vee \beta))}_{\text{No}}$$

$$?\Diamond(\alpha \vee \beta) \equiv \underbrace{\Diamond(\alpha \vee \beta)}_{\text{Yes}} \mathbin{\vee\!\!\!\vee} \underbrace{\neg\Diamond(\alpha \vee \beta)}_{\text{No}}$$

Therefore a simple extension of the scalar approach predicts one of the following patterns:

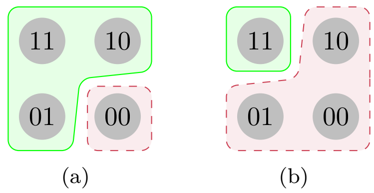


Figure 4: Response patterns predicted by a simple inquisitive extension of the scalar approach when the question radical is not exhaustified (left) and when it is exhaustified (right).

Scalar accounts have no trouble with the asymmetry between *FC* and *DP*, however, they are committed to the claim that scalar reasoning is involved in the process of computing the *FC* inference. For instance, scalar reasoning causes the *delay effect*, i.e. it takes people longer to compute a scalar implicature than computing to access the literal meaning of the same sentence (e.g. Bott and Noveck, 2004; Bott et al., 2012). Chemla and Bott (2014) empirically showed that the predictions of scalar theories regarding the processing of *free choice* sentences are incorrect. Tieu et al. (2019) as well as Marty et al. (2021) confirmed that for other related inferences such as *dual prohibition* or *negative free choice.*

Predictions regarding the processing of *FCQ*s follow from the results of Bott and Noveck (2004). We know that computing scalar implicatures takes longer than accessing the literal meaning. Scalar reasoning is needed only in the *positive* case (*FC* or the *Yes* response); in the negative case (*DP* or *No*), computing the literal meaning only yields the correct interpretation. Therefore, scalar theories predict a delay effect for the reasoning about the *Yes* response particle and no effect on *No.* Moreover, complicated scalar reasoning may take longer than literal-meaning reasoning.

2.5 Homogeneity

The core idea of the homogeneity approach to *free choice* is that "*disjunctions are homogeneous with respect to modal status...*"(Goldstein, 2019, p.35). This means that disjunction under a deontic modality can only be meaningful/assertible if the disjuncts have the same truth value; if they are either *both* true or *both* false. In Goldstein's account, homogeneity is a semantic presupposition.[13]

To model homogeneity Goldstein (2019) proposes to make use of trivalent logic. If a formula does not satisfy homogeneity, it will have the third truth value – *undefined.* If we take formulas to be assertions, then we can easily interpret *undefined* as unassertable (e.g. because of pressuposition failure). Goldstein proposes two logical frameworks:

- Homogeneous Alternative Semantics, where $\Diamond\varphi$ is defined only if all the alternatives in $[\![\varphi]\!]$ *have the same truth value,* where $[\![\alpha \vee \beta]\!] = [\![\alpha]\!] \cup [\![\beta]\!]$.
- Homogeneous Dynamic Semantics, where $\alpha \vee \beta$ is defined only if either both $\Diamond\alpha$ and $\Diamond\beta$ or both $\neg\Diamond\alpha$ and $\neg\Diamond\beta$ are supported.

semantics, it is also possible for the exhaustivity operator to scope over the question: $Exh(?\Diamond(\alpha \vee \beta))$. However, it is unclear what the set of alternatives would be generated by an inquisitive prejacent. We leave it to the proponents of this approach to figure out if this is a possible way to solve the issue with *FCQ*s.

[13]Bar-Lev (2018) treats homogeneity as an implicature and Del Pinal et al. (2023) include it as a formal tool in the exhaustification algorithm. The proposal here is to treat homogeneity as a primary notion – independent of scalarity.

To account for *FC* it is crucial to observe that its premise ($\Diamond(\alpha \vee \beta)$) in classical logic implies that at least one of the disjuncts is permitted. By homogeneity, we know that they need to be both allowed or both not allowed. Therefore, they must be both permitted. *DP* is a classically valid inference, and homogeneity does not affect it. Its premise ($\neg\Diamond(\alpha \vee \beta)$) implies that both disjuncts are not permitted, which does not violate homogeneity.

The core idea about the homogeneous behaviour of disjuncts under a deontic modality is easily extrapolated to questions. If either both choices are allowed or both are not allowed, then the positive part of an *FCQ* must refer to the case where both are allowed and the negative to the case where neither is allowed. According to the formalism proposed by Goldstein (2019), homogeneity is applied on the level of each part of the question, following the reasonings about *FC* and *DP* from the declarative case. However, a more general pragmatic theory would also work accordingly with the aforementioned intuition.

$$?\Diamond(\alpha \vee \beta) \equiv \underbrace{\Diamond(\alpha \vee \beta)}_{\text{Yes}} \mathbin{\vee\!\!\!\vee} \underbrace{\neg\Diamond(\alpha \vee \beta)}_{\text{No}}$$

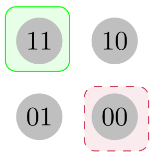


Figure 5: Response patterns predicted by Homogeneity.

Regarding processing, the homogeneity approach makes predictions opposite to the scalar theories. Performing a (local) pragmatic weakening, e.g. suspending presupposition (Schwarz, 2013) takes longer than computing the meaning using the pragmatic effect. This is known as *reversed delay effect* and is empirically established in various Free Choice inferences (Chemla and Bott, 2014; Tieu et al., 2019). Moreover, this analysis suggests that processing the contexts where the presupposition is not satisfied should take longer.

2.6 Neglect-zero

Aloni (2022) proposed a high-level pragmatic tendency of human reasoning called *neglect-zero*. She postulates that people consistently neglect empty configurations in reasoning and have trouble accommodating them if necessary to evaluate a sentence. This cognitive bias is independently motivated by studies about quantifiers by (Bott et al., 2019) and Ramotowska et al. (2022).[14]

To model linguistic behaviour Aloni (2022) proposes Bilateral State-based Modal Logic (BSML). In BSML the formulas are interpreted with respect to a state (set of possible worlds), rather than with respect to a single world. Moreover, BSML defines support ($\models$) and anti-support ($=\!\mid$) conditions for formulas to capture their assertability and rejectability. Negation in BSML is bilateral:

$M, s \models \neg\varphi$ iff $M, s =\!\mid \varphi$.

$M, s =\!\mid \neg\varphi$ iff $M, s \models \varphi$.

As we can observe from the semantic clauses for negation, sentences which are neither assertable nor rejectable are unassertable. Aloni defines disjunction as a *split* of state into two parts, where each part supports one of the disjuncts. Note that a part can in principle be empty. Modality is defined in a standard state-based fashion (Humberstone, 1981).

$M, s \models \varphi \vee \psi$ iff $\exists t, t' : t \cup t' = s$ and $M, t \models \varphi$ and $M, t' \models \psi$.

$M, s =\!\mid \varphi \vee \psi$ iff $M, s =\!\mid \varphi$ and $M, s =\!\mid \psi$.

$M, s \models \Diamond\varphi$ iff $\forall w \in s\ \exists t \subseteq R[w] : t \neq \emptyset :$ and $M, t \models \varphi$.

$M, s =\!\mid \Diamond\varphi$ iff $\forall w \in s : M, R[w] =\!\mid \varphi$.

[14]To learn more about *neglect-zero* please consult the paper by Aloni (2022).

Observe that these definitions are equivalent to the classical semantic clause for disjunction and existential modal. The key non-classical component of this approach is the *non-emptyness atom* with support and anti-support conditions as follows:

$M, s \models$ NE iff $s \neq \emptyset$.

$M, s \dashv$ NE iff $s = \emptyset$.

Without NE, BSML is just classical logic. NE introduces non-classicality by enforcing the omission of empty configurations in the evaluation of (parts of) sentences. For instance the simple state $s = \{w_p\}$ where $w_p \models p$ and $w_p \not\models q$ supports $p \vee q$, as we can split it into $t = s$, which supports p and $t' = \emptyset$, which supports q. However, s does not support $p \vee (q \wedge \text{NE})$, since the second disjunct can neither be supported by the empty state nor by a non-empty subset of s. Observe that $p \vee (q \wedge \text{NE})$ is also not rejected by s, since p is not rejected.[15] Therefore this formula is not *false* at s, it is just unassertable.

Aloni (2022) postulates, that conversational reasoning enforces pragmatic enrichment of sentences. We denote a pragmatically enriched formula using $[\cdot]^+$. Enrichment ensures that no part of a sentence is supported by an empty configuration by recursively adding NE as a conjunct over sub-formulas. The enriched premise of *FC* looks as follows:

$[\Diamond(\alpha \vee \beta)]^+ = \Diamond((\alpha \wedge \text{NE}) \vee (\beta \wedge \text{NE}))$

Let's see how BSML can accommodate *free choice* and *dual prohibition*. Recall the formulations of *FC* and *DP* from page 1. Let's show that they both hold in BSML in the enriched version, i.e. that $[\Diamond(\alpha \vee \beta)]^+ \models \Diamond\alpha$ (*FC*) and $[\neg\Diamond(\alpha \vee \beta)]^+ \models \neg\Diamond\alpha$ (*DP*).

Proof. FC: Suppose $M, s \models [\Diamond(\alpha \vee \beta)]^+$ Then $\forall w \in s\ \exists t \subseteq R[w] : t \neq \emptyset :$ and $M, t \models (\alpha \wedge \text{NE}) \vee (\beta \wedge \text{NE})$. Therefore $\exists t', t'' : t' \cup t'' = R[w]$ and $M, t' \models \alpha$ and $M, t'' \models \beta$. Where $t', t'' \neq \emptyset$. Since $t' \subseteq t \subseteq R[w]$: $\forall w \in s\ \exists t' \subseteq R[w] : t' \neq \emptyset :$ and $M, t' \models \alpha$. Thus $M, s \models \Diamond\alpha$. □

Proof. DP: $M, s \models [\neg\Diamond(\alpha \vee \beta)]^+$, then $\forall w \in s : M, R[w] \dashv (\alpha \wedge \text{NE}) \vee (\beta \wedge \text{NE})$, so $\forall w \in s : M, R[w] \dashv (\alpha \wedge \text{NE})$. Since $R[w] \neq \emptyset$ then $M, s \models \neg\Diamond\alpha$. □

An extension of BSML with a mechanism which allows for the modelling of questions is independently motivated by the formal properties of this extended system (Anttila, 2021). This system adapts the following support and anti-support clauses for the inquisitive disjunction:

$M, s \models \varphi \mathbin{\vee\!\!\!\vee} \psi$ iff $M, s \models \varphi$ or $M, s \models \psi$.

$M, s \dashv \varphi \mathbin{\vee\!\!\!\vee} \psi$ iff $M, s \dashv \varphi$ and $M, s \dashv \psi$.

Pragmatic enrichment is again defined over the subformulas of the extended system. In BSML with inquisitive disjunction, the positive part of a Free Choice Question is just the premise of *FC* and the negative part is the premise of *DP*. Therefore the question is supported if and only if either *FC* or *DP* is supported, which means that, by the proofs in the previous section, *Yes* corresponds to the case where both choices are allowed and *No* to the case where neither is:

$$[?\Diamond(\alpha \vee \beta)]^+ \equiv \underbrace{[\Diamond(\alpha \vee \beta)]^+}_{\text{Yes}} \mathbin{\vee\!\!\!\vee} \underbrace{[\neg\Diamond(\alpha \vee \beta)]^+}_{\text{No}} \wedge \text{NE}$$

BSML resolves the issue of *free choice* by postulating a high-level pragmatic principle governing conversations in natural language. The interesting prediction made by Aloni (2022) is that this principle can be suspended locally (for the sake of processing a single utterance) or globally (for the entire context/conversation). Global suspension is possible in, e.g. mathematical or logical discourse, where we choose to interpret disjunction using the classical truth tables. Global suspension would yield the following (classical) interpretation of *FCQ*s:

$$?\Diamond(\alpha \vee \beta) \equiv \underbrace{\Diamond(\alpha \vee \beta)}_{\text{Yes}} \mathbin{\vee\!\!\!\vee} \underbrace{\neg\Diamond(\alpha \vee \beta)}_{\text{No}}$$

[15]For the full semantics of BSML, and other ways of modelling *neglect-zero* consult the paper by Aloni (2022)

For instance, on an algebra exam the question (11) should be answered with *Yes* even though this answer seems counter-intuitive in a regular conversation.

(11) May $n + 3$ be grater or equal to 2?

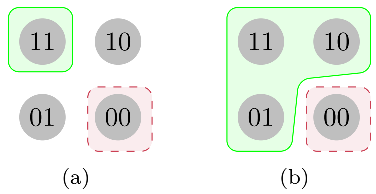


Figure 6: Response patterns predicted by BSML with neglect-zero: default interpretation without suspension (left) and with suspension (right).

Similarly to *homogeneity*, the *neglect-zero* theory makes opposite predictions regarding processing to the scalar theories. As indicated by Aloni (2022) local suspension (e.g. for a single assertion in a normal conversation) is more costly and causes more difficulty in processing than computing the meaning using the pragmatic effect (*reversed delay effect*) (Bott et al., 2019; Ramotowska et al., 2022). Global suspension may be difficult to acquire, but once it is in place, it does not influence the processing difficulty.

Moreover, Bott et al. (2019) as well as Ramotowska et al. (2022) showed that it *takes longer to process zero-models* than non-zero-models, at least in the domain of quantifiers. This is similar to the results regarding the processing of contexts which violate the homogeneity presupposition. Thus the non-scalar theories (*homogeneity* and *neglect-zero*) make the same predictions except that it is unlikely that a semantic presupposition could be easily (globally or locally) suspended, while *neglect-zero* is suspendable.

The next section will report on an experiment testing the predictions of all the above theories.

3 Experiment

This experiment investigates the conversational relation between Free Choice Questions and the response particles *Yes* and *No*. More specifiaclly it addresses the following research questions: What do response particles correspond to as replies to an FCQ? What is the source (semantic/pragmatic) of the inferences triggered by the response particles? Answering these questions will allow us to evaluate the predictions of the theories of declarative *free choice* discussed in Section 2.

3.1 Methods

To investigate these issues, we adapted the sentence-picture acceptability task used by e.g. Marty et al. (2021), to a *conversation-picture acceptability task*, in which participants are presented with a short conversation consisting of a question and an answer as well as, with a picture representing what is allowed and not allowed in a given context. The participants were asked to evaluate the answer with respect to the picture. We collected participants' answers, as well as their reaction times. Figure 3.1 showcases a trial of the experiment.[16]

3.1.1 Participants

Using *prolific.co* we recruited 60 native speakers of English located in the UK or in the US to participate in our experiment. Participants were informed about their rights and that the study was approved by the Ethics Committee of the Faculty of Humanities of the University of Amsterdam (FGW-341). The participants were paid £2.25 for their participation.

[16] A demo version of this experiment is available at: `https://www.tklochowicz.com/experiment_FCQ`

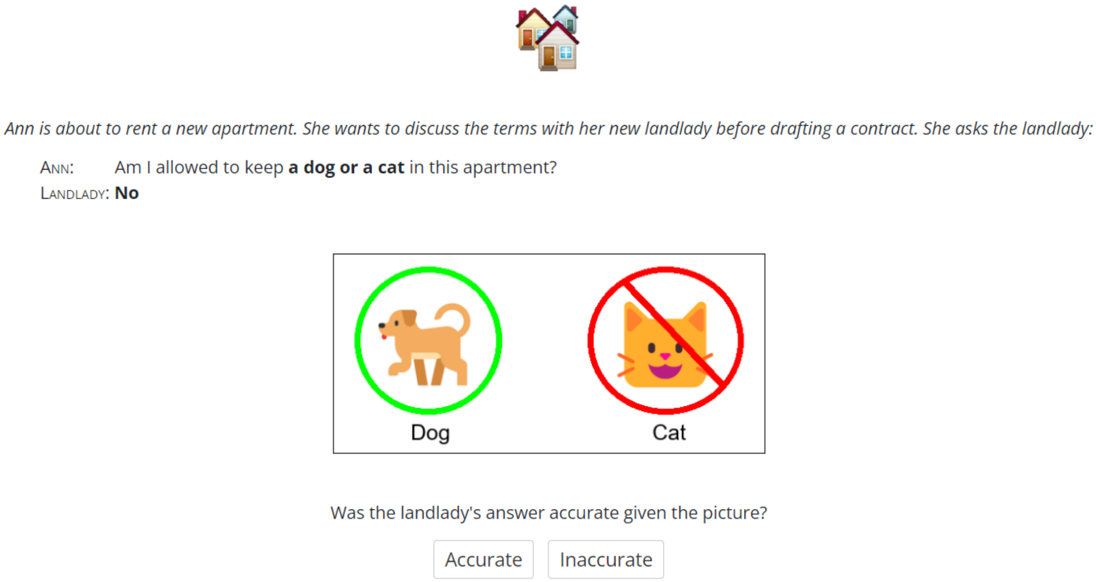
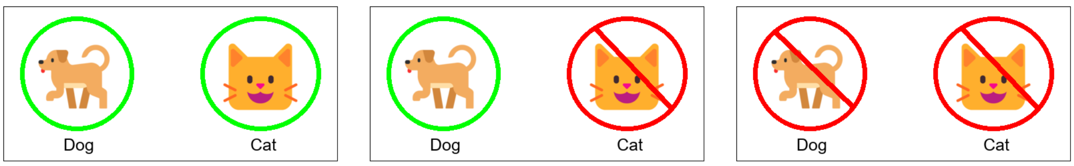


Figure 7: An example of a target trial item.

3.1.2 Study design

Free Choice Questions involve two items separated by disjunction. Each item can be allowed or not allowed in a given context (represented by a picture). Therefore each tested *FCQ* was evaluated with respect to three contexts (*both allowed*, *one allowed*, *neither allowed*):



(a) *both allowed* (control) (b) *one allowed* (target) (c) *neither allowed* (control)

Figure 8: Contexts tested in the experiment.

For each context, we considered two response particles (answers): *Yes* and *No*. We created two scenarios to investigate the difference between two types of speech acts i.e. *granting permission* (12) and *reporting permission* (13). Thus, our design was $3 \times 2 \times 2$ (*contexts* $\times$ *response particles* $\times$ *scenarios*).

(12) *Ann is about to rent a new apartment. She wants to discuss the terms with her new landlady before drafting a contract. She asks the landlady:*

(13) *Bill is in London at a tourist office. He wants to know more about the tourist pass they offer. He asks the employee of the office:*

Therefore for each response particle, we had True, False and Target conditions. For the particle *Yes* the True condition was *both allowed* and False was *neither allowed*. For *No* it was the other way around: True was *neither allowed* and False was *both allowed*. The Target conditions for both particles were the *one allowed* contexts.

3.1.3 Materials

For each scenario, we created four pairs of items which were used to create the pictures representing the contexts. For scenario (12) the pairs consisted of animals that Ann would like to (potentially) keep at her apartment, and for scenario (13) of tourist attractions in London, which Bill would like to visit with the tourist pass.[17].

A trial of the experiment consisted of a scenario represented both graphically and textually, a Free Choice Question asked by the main character (Ann or Bill) and an answer given by the second character (the landlady or the employee of the tourist office). We used a squared picture with a pair of items to represent the context in each trial. Each item was either in a green or a red crossed-out circle. The green circle represented that the item is allowed and the red circle that it is not. In each trial, the participants were asked to evaluate the answer of the second character given the picture using two buttons labelled: ACCURATE, INACCURATE.

Filler trials were very similar to the test trials, but instead of a Free Choice Question, the main character asked a polar question about one of the items like: *May I keep a dog in this apartment?*.

3.1.4 Procedure

The experiment was designed using the jsPsych library by de Leeuw et al. (2023), and it was published using *cognition.run*.

In the beginning, the participants were presented with information about the experiment and prompted to express their informed consent for participation. Afterwards, they were asked to confirm that they were a native speaker of English.

To familiarise the participants with the layout of the experiment and the interpretation of its parts, especially the pictorial representation of the context and the scenarios, we started the experiment with a training phase consisting of filler items. For each answer in the training phase, the participants received feedback. If the answer was correct the word "*Correct!*" appeared briefly on the screen. If the answer was incorrect, the word "*Incorrect!*" appeared alongside the information that they had to wait 4 seconds and try the training trial again. The participant could not continue with the experiment without getting all the training items correctly.

In the phase of design, we noticed that the real-life contexts may cause participants to evaluate not only the consistency or correctness of the second character's answer but also their politeness, precision or work skills. Since these readings of the experimental task could skew the results, we prepared training items, which would prevent them. For instance, in the context where the landlady wants to allow for a dog, but not for a cat, her answer "*No*" to the question "*May I keep a dog in this apartment?*" had to be evaluated as accurate.

After the training phase, the participants were prompted that they would not receive feedback anymore, and the test trials would begin. In total there were 8 pairs of items, each was presented in three contexts and with two polarity particles. Therefore the participants had to answer $8 \times 3 \times 2 = 48$ test trials. Moreover, to check if the participants understood the task and were paying attention we used 24 filler items. Those 72 trials were presented in random order. Before each trial, a fixation cross would appear for a random amount of time (between 0.25 and 2 seconds) to allow for a more accurate measurement of reaction times.

3.1.5 Data treatment and statistical analysis

The data was collected using *cognition.run* and downloaded as a JSON file. We wrote a simple *python* program to extract the information which was important for the study and convert the files into a *.csv*, which is more sustainable for data storage and processing.

To perform the statistical analysis we used language$R$, the *lme*4 library and other libraries as well as R-studio IDE (R Core Team, 2021; Bates et al., 2015; RStudio Team, 2020). We used *mixed*

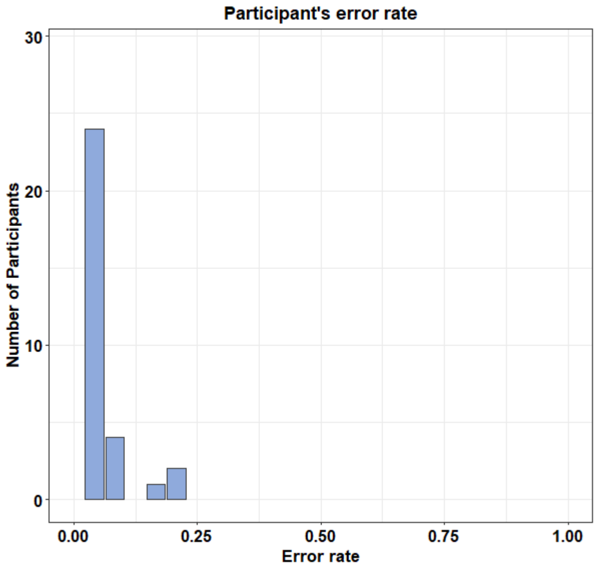
[17]Pairs of items for (12): (*a dog, a cat*), (*a lizard, a parrot*), (*a mouse, a hamster*), (*a snake, a spider*); Pairs of items for (13): (*Big Ben, the London Eye*), (*Buckingham Palace, Westminster Abbey*), (*the National Gallery, the British Museum*), (*Tower Bridge, the London Dungeon*)

logistic regression as a model for acceptance rates and *linear regression* for reaction times. For data visualisation we used *ggplot2* by Wickham (2016).[18]

3.2 Results

3.2.1 Filler items

We measured each participant's error rate on filler items, to check if they understood the task, and paid attention throughout the study. The distribution of error rate is represented in Figure 3.2.2. We had two participants with $\approx 21\%$ error rate and one with $\approx 17\%$. The rest did much better ($<$ 5% error rate). Overall, the mean error rate was 3.2%. Since the error rates were within standard norms, we decided not to exclude any participants from the analysis. A small error rate on filler trials indicates that the task was simple and understandable for participants. This conclusion is further confirmed by the strong acceptability of TRUE conditions and strong rejection of FALSE conditions (See below).



3.2.2 Acceptance rate

Figure 9 presents the acceptance rates for the six conditions. We used *mixed logistic regression* to model the differences between the 6 conditions (3 *contexts* $\times$ 2 *response particles*) in pairwise comparisons. For both response particles, the TRUE conditions were almost uniformly accepted, and the FALSE conditions were uniformly rejected. The two control conditions were significantly different from each other ($|\beta| > 4$, $p < 0.001$). There is no significant difference between the particles *Yes* and *No* on their respective TRUE and FALSE conditions (all $|\beta| < 1$ $p > 0.1$).

For the *Yes* particle there are significant differences ($p < 0.001$) between the TRUE ($\beta \approx 5.6; 99\%$) and the TARGET condition ($\beta = -7.1; 18\%$) as well as between the FALSE ($\beta \approx -4.8; 0.9\%$) condition and the TARGET ($\beta \approx 3.3; 18\%$)). However, for each particle, the TARGET condition is still significantly closer (more similar) to the FALSE condition than to the TRUE condition ($p < 0.001$). To test that we compared the TARGET condition to the mean of TRUE and FALSE using contrast coding.

Similarly, for the *No* particle, there is a significant ($p < 0.001$) difference between the TRUE ($\beta \approx 3.9; 98\%$) and the TARGET condition ($\beta = -6.2; 9\%$) as well as between the FALSE ($\beta \approx -4.2; 1.5\%$) condition and the TARGET ($\beta \approx 1.1, p < 0.001$)).[19] Again, the TARGET condition is significantly

[18] The data, the code and the statistical analysis are available at `https://uvaauas.figshare.com/projects/Free_Choice_Questions/177987`.

[19] In this condition the mixed model was not converging, therefore we used simple logistic regression without the mixed effects.

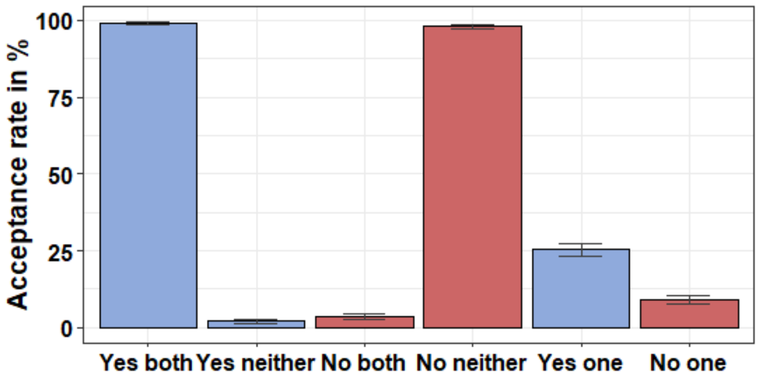


Figure 9: Acceptance rates of the six conditions

closer (more similar) to the FALSE condition than to the TRUE condition ($p < 0.001$).

Moreover, the difference between the TARGET for the *Yes* particle and the TARGET for the *No* particle is significant ($p < 0.001$). We found significant interaction between conditions (FALSE, TARGET) and response particles ($p < 0.001$). The interaction is presented in Figure 10. Thus, even though both target conditions are closer to FALSE than to TRUE, the TARGET condition for the *No* particle is significantly closer to its FALSE than the TARGET condition for the *Yes* particle to its FALSE condition.

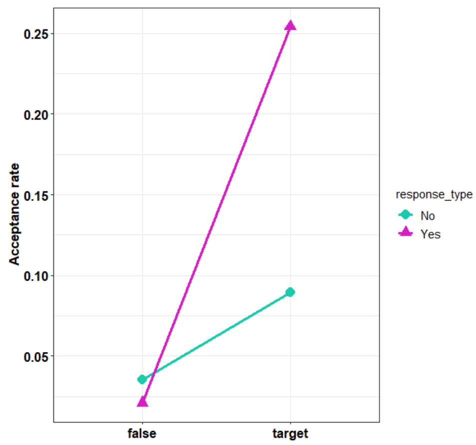


Figure 10: Interaction between conditions and response particles

We did not find differences between the complete analysis and separate analyses of the scenarios involving the landlady and the scenario about the tourist office (See above). The only one is that in the tourist office scenario, the difference between the TARGET for the *No* particle and its FALSE condition is roughly the same in size, but no longer significant ($p > 0.05$). Moreover, we did not find any significant differences in the direct comparison of the two scenarios (all $p > 0.1$).

winning-participant

Figure 11 represents the distribution of the participants' mean acceptance rate of the 6 conditions. The bold line indicates the median approval rate. We can see that there are few participants with a mean acceptance rate much lower than 100% for TRUE conditions and much higher than 0% for FALSE conditions. Moreover, the median acceptance rates for all conditions are 1 or 0. Again, the fact that the median acceptance rate for the TARGETs is 0 indicates that they are more similar to FALSE than to TRUE. However, we can observe that there were participants who consistently judged the conversation in *Yes one allowed* to be accurate. We will call them the *non-FC participants*, as they did not compute the *FC* inference. We can observe that these participants should not be considered outliers, but a subgroup of participants. These participants are also clearly visible in Figure 12.

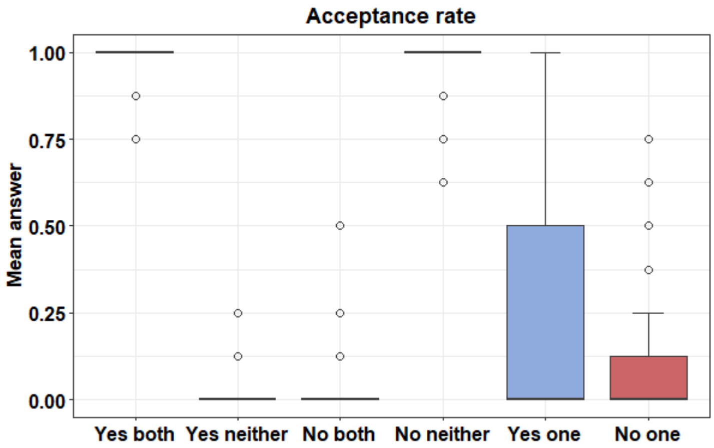


Figure 11: Mean acceptance rates of participants

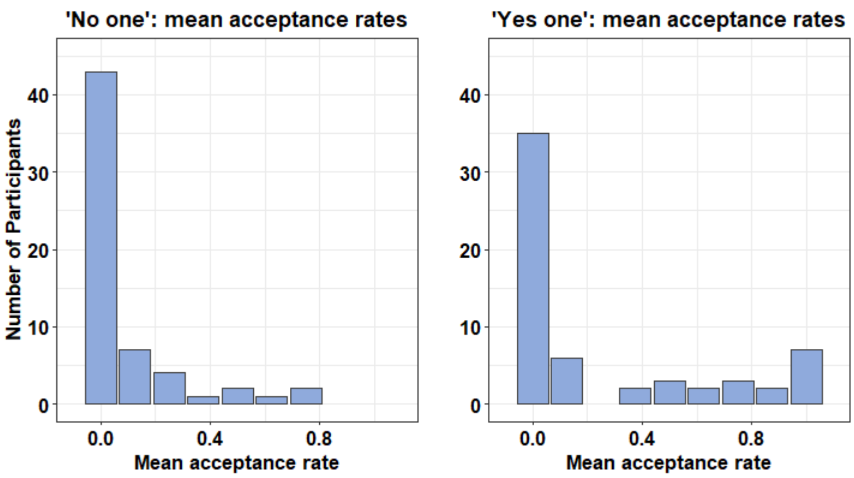


Figure 12: Mean acceptance rates of targets per participant

3.2.3 Reaction times

Reaction times were collected using the tools built into jsPsych. The mean reaction time to a trial was 4.8 seconds with a standard deviation of approximately 7.75 seconds. We removed 24 outliers which lay further than 3 standard deviations from the mean (which took longer than 27 seconds) leaving us with 2856 trials with a mean reaction time of 4.3 seconds and a standard deviation of 3 seconds. The data is represented in Figure 13.

In the collected reaction times, we observed two main effects: 1. Trials where the *No* particle was used as the response to an *FCQ*, took significantly longer to evaluate than those with *Yes* ($\beta \approx 0.38$sec, $p < 0.001$). 2. The *target* contexts took significantly longer to evaluate ($\beta \approx 1.3$sec, $p < 0.001$) than the controls. These effects are displayed in Figure 14.

These effects are cumulative. The highest mean reaction time is found for TARGET trials involving the particle "No". On average they take approximately 0.5 sec longer than TARGET trials with the "Yes" ($p < 0.01$). People needed on average around $1sec$ less to complete the control trials with the "No" particle, but there was no significant difference between these two ($p > 0.7$). Moreover, people needed approximately $1.3sec$ more in the control trials with the "Yes" particle (both $p < 0.001$),

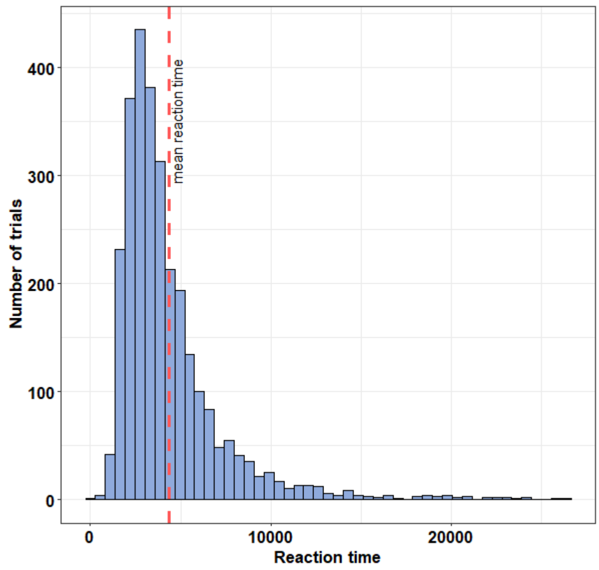


Figure 13: Distribution of reaction times over trials.

and there was no significant difference between these two controls either (p>0.06). The difference between the controls with the "No" particle and the "Yes" particle was significant ($\beta \approx 0.3sec$, $p < 0.01$).

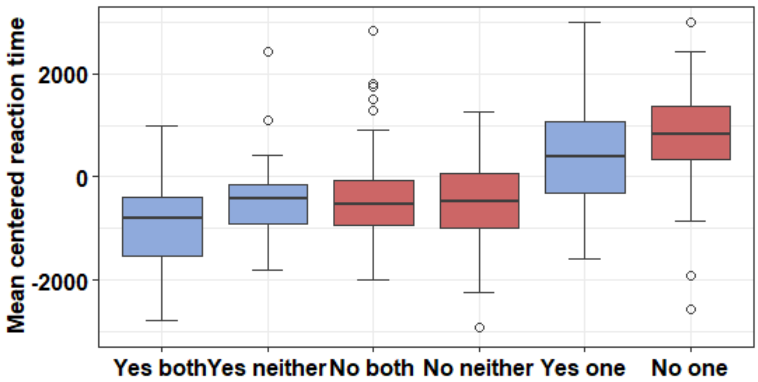


Figure 14: Centered mean reaction times per condition

To ensure that the target effect is not due to the visual difference between targets and controls, we performed the same analysis on the filler items. They are visually the same as the tests but contain only one item in the question instead of the disjunction (e.g. "May I keep a dog?"). We found a similar negation effect ($\beta \approx 0.2$sec, $p < 0.001$), but the target effect ($\beta \approx 0.3$sec, $p < 0.001$) was significantly smaller: The interactions between the context type (*both allowed*/*one allowed* as well as *neither*/*one allowed*) and item type (*filler*/*test*) was significant (see Figure 15: $\beta \approx 1.1$sec, $p < 0.001$)).

3.2.4 Delay effect

We did not observe any delay effect: in the *target* context and the response particle *Yes*, it took participants as long to accept as to reject. The analysis of the particle *No* was inconclusive, as there is little acceptance data (only 9%), but we observed an insignificant tendency that accepting takes longer (see Figure 16).

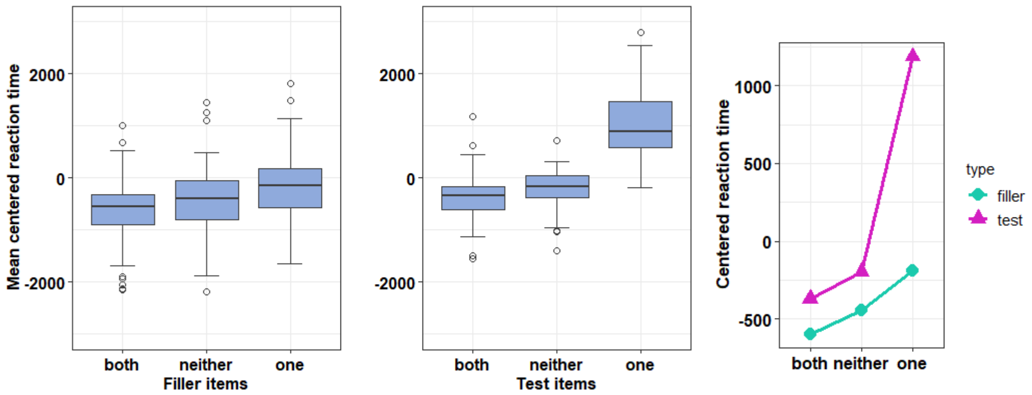


Figure 15: Reaction times by context on test and filler items and interaction between context and filler/test.

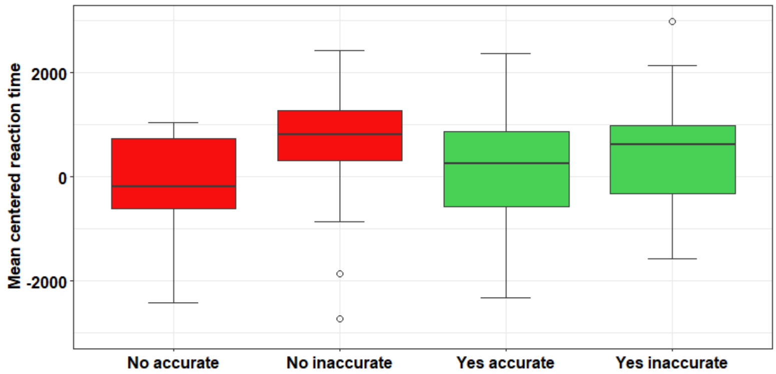


Figure 16: Comparison of the reaction times on TARGET items by response particle and the answer given by participant (*accprate/inaccurate*).

3.3 Discussion

Good performance on filler items and robust results on the controls indicate that the participants understood the task and that we managed to prevent politeness or precision readings. In particular, we ruled out the possibility that a short answer like *Yes* is evaluated as inaccurate because it is not polite enough to use as a response to a client's question. This fact allows us to draw meaningful conclusions from the collected data. First, recall the possible answer patterns:

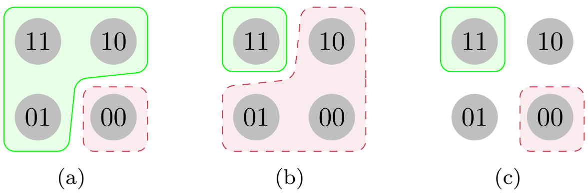


Figure 17: Possible patterns for response particles.

In the experiment, we established that the response particle *Yes* as an answer to *FCQ* corresponds to *both allowed* and the response particle *No* to *neither allowed*. Therefore, the correct pattern of responses is represented by Figure 17(c). Thus, we showed that *FCQ*s pattern with declarative *free choice* and *dual prohibition*. Moreover, we observed that the acceptance rates for the *target* conditions are lower than for the (literal meaning) controls, which suggests that they are weaker, more difficult, or have a pragmatic source.

Moreover, we found that there is a group of participants (*non-FC participants*) who consistently indicated that *Yes* corresponds to the classically valid *at least one allowed*. Their answers are represented in Figure 17(a).[20]

We found that the *one allowed* contexts make processing of *FCQ*s more difficult, but they do not increase the difficulty of the processing of non-disjunctive polar questions with the deontic operator. This suggests that some pragmatic principles of utterance regarding *FCQ*s and the response particles are violated in those cases. We also found that the response particle *No* is more difficult to process then the particle *Yes*. We called this effect the *negation effect*.

Analysing the *one allowed* contexts, we found no delay effect. The participants were as quick to accept as to reject. The tendencies point towards the *reversed* delay effect, but the differences were not significant.

4 General Discussion

The theories by Goldstein (2019) and Aloni (2022) correctly predict the meaning of response particles as answers to Free Choice Questions, since it follows from them that *Yes* corresponds to *both allowed* and *No* to *neither allowed*. Semantic and scalar theories will have to do some work to accommodate this data.

The issue of Deontic Inquisitive Logic regarding *DP* can be resolved by, e.g. accepting bilateral negation.[21] If the universal quantifier at the beginning of the definition of the diamond remains universal in the rejection clause, DIL can account for *DP*:

$M, s \models \neg\Diamond\varphi$ iff *for all* $w \in s$, for all $Y \in ALT_M(\varphi) : Y \cap R[w] = \emptyset$

This would also change the prediction of this theory regarding *FCQ*, as they would match the results of the experiment. However, it is unclear what would be a result of the complete bilateralisation of inquisitive logic. We will not dive deeper into this issue and just accept this proposal as a possible solution to the *free choice* puzzle, as far as the interpretation of *Yes* and *No* is concerned. However, purely semantic solutions will always have trouble accommodating the processing data regarding both declarative and inquisitive versions of *free choice*.

Proponents of the exhaustification approach could develop a theory of questions that would apply exhaustification to each classically derived answer.[22] In the case of *FCQ*, this would yield the question to consist of the exhaustified premise of *FC* and the exhaustified premise of *DP*, which are accommodated in the declarative versions. However, it is unclear what such a theory would mean to standard questions and whether it is indeed true that every answer to any question should always be exhaustified. Moreover, it is unclear how such an account could accommodate for embedded *FCQ*s, where no answer is uttered in a conversation, as in (14):

(14) Mary knows whether Ann may keep a dog or a crocodile.

Moreover, Aloni (2022) can explain the behaviour of the *non-FC participants* and (partially) the difference between *Yes* and *No* on target items through global suspension of neglect-zero. We would expect that the homogeneity presupposition cannot be globally "suspended". To enable suspension, homogeneity could be interpreted as a more high-level pragmatic principle which is independently motivated and accounted for, e.g. in plural definites (Križ, 2019). Global suspension of neglect-zero should not cause any difference in reaction times, which is confirmed in the data. The scalar theories could explain the non-FC participants as those who do not compute implicatures.

[20] Note that we did not find differences between the two scenarios, which were standing for different speech acts (granting vs. reporting permission). However, in the conversation-picture acceptability task, it isn't easy to ensure that the participants will see the difference between the two speech acts. In the real conversations involving granting permission, we assume that the person who grants it has full freedom of doing so. The pictures in the task seem to suggest that they have some pre-established rules limiting them. We decided that having a similar method and comparable results to other research in this area (e.g. Bott et al. (2019); Marty et al. (2021); Ramotowska et al. (2022)) is more important than ensuring that we observe a difference here. Our other results are not affected by this choice. A different task should be designed to see whether these two speech acts differ. It could involve production or priming. We are leaving that for further research.

[21] For a semantic solution along those lines, see Willer (2018).

[22] Another possible approach would be to adopt a version of presuppositional exhaustification along the lines of Del Pinal et al. (2023). This idea will be discussed in more lnght in the final version of this paper.

However, we would expect them to be quicker than those who compute them, which is not what we found.

Furthermore, the non-scalar theories correctly predict longer reaction times for *target* contexts, which violate the homogeneity presupposition and are zero models for disjunction. This prediction is unavailable for the scalar approaches since scalar reasoning is needed in *one allowed* contexts only with the *Yes* response particle. Proponents of the scalar theories could argue that the longer reaction times can be explained by the need for disambiguation between exhaustified and non-exhaustified formulations, as they differ in meaning only in the *one allowed* contexts. However, we would expect that this effect and the scalar effect should behave cumulatively, i.e., that *one allowed* contexts take longer than the others, but the positive (*Yes*) takes longer than the negative (*No*). We found that the negative takes longer. Since the semantic approaches predict uniform reaction times across all contexts, they will have trouble accommodating the data.

We did not find any delay effects predicted by theories of *free choice*, but our analysis was difficult due to low acceptance rates of the *one allowed* contexts. It is worth noticing that scalar inferences differ in behaviour depending on various factors (Van Tiel et al., 2016). Therefore, simple extrapolation of predictions from declaratives to questions may not be accurate. As far as we are concerned, there is no systematic study of delay effects in questions containing implicatures, presuppositions and other pragmatic inferences. To fully evaluate the results regarding the delay effect, we would need to compare them with such a study.

5 Conclusion

We performed an experiment to test the predictions of various theories explaining declarative *free choice* inferences regarding the inquisitive version of those inferences *Free Choice Questions.* We proposed extensions of those theories and argued that the predictions should hold independently from the way these theories are extended to capture questions. In the experiment, we tested participants' intuitions about the meaning of response particles (*Yes* and *No*) as well as the way they process these questions. The processing was analysed through the study of reaction times.

We found that *Yes* corresponds to *free choice* and *No* to *dual prohibition*, i.e. *Yes* to the case where both choices are allowed and *No* to the case where neither is allowed. These results correspond to declarative *free choice* and *dual prohibition.* Moreover, the data regarding the processing of those questions is more similar to the predictions made by non-scalar pragmatic theories than to scalar ones. The collected data poses a challenge to semantic and scalar approaches to free choice and supports non-scalar pragmatics as a uniform solution to the free choice puzzle.

As further research, we would like to study other possible answers to Free Choice Questions (e.g. involving cancellation) and embedded Free Choice Questions: *Mary knows whether Bill may go to the park or to the beach.* Moreover, we would like to compare the results regarding *FCQ*s to Scalar Questions (e.g.*Did some students pass the exam?*) and Homogeneity Questions (e.g. *Did the boys go to the park?*). Moreover, we would like to investigate further the *negation effect* associated with response particle *No* and compare it to the processing of sentential negation.

References

Aher, M. (2013). *Modals in Legal Language.* PhD thesis, Institute of Cognitive Science,University of Osnabrück.

Aloni, M. (2007). Free choice, modals, and imperatives. *Natural Language Semantics*, 15:65–94.

Aloni, M. (2016). Fc disjunction in state-based semantics. *Talk given at Logical Aspects of Computational Linguistics.*

Aloni, M. (2018). Fc disjunction in state-based semantics. *Unpublished manuscript, available at https://www.marialoni.org/resources/Aloni2018.pdf.*

Aloni, M. (2022). Logic and conversation: the case of free choice. *Semantics and Pragmatics*, 15:5–EA.

Alonso-Ovalle, L. (2006). *Disjunction in alternative semantics.* University of Massachusetts Amherst.

Anttila, A. (2021). *The Logic of Free Choice. Axiomatizations of State-based Modal Logics.* PhD thesis, Masters thesis, Institute for Logic, Language and Computation.

Bar-Lev, M. E. (2018). *Free choice, homogeneity, and innocent inclusion.* PhD thesis, Hebrew University of Jerusalem dissertation.

Bar-Lev, M. E. and Fox, D. (2020). Free choice, simplification, and innocent inclusion. *Natural Language Semantics*, 28(3):175–223.

Barker, C. (2010). Free choice permission as resource-sensitive reasoning. *Semantics and Pragmatics*, 3:1–38.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bott, L., Bailey, T. M., and Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, 66(1):123–142.

Bott, L. and Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, 51(3):437–457.

Bott, O., Schlotterbeck, F., and Klein, U. (2019). Empty-set effects in quantifier interpretation. *Journal of Semantics*, 36(1):99–163.

Chemla, E. and Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition*, 130(3):380–396.

Ciardelli, I., Groenendijk, J., and Roelofsen, F. (2018). *Inquisitive semantics.* Oxford University Press.

Ciardelli, I. A. (2016). *Questions in logic.* PhD thesis, University of Amsterdam.

Cremers, A., Hoeks, M., Lisowski, G., and Pesetsky, J. (2017). Experimental evidence for a semantic account of free choice. In *Colloque de Syntaxe et Semantique à Paris.*

Dayal, V. (1996). *Locality in WH Quantification. Questions and Relative Clauses in Hindi.* Springer Dordrecht.

de Leeuw, J. R., Gilbert, R. A., and Luchterhandt, B. (2023). jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, 8(85):5351.

Del Pinal, G., Bassi, I., and Sauerland, U. (2023). Free choice and presuppositional exhaustification. *unpublished manuscript lingbuzz/006122.*

Fox, D. (2007). Free choice and the theory of scalar implicatures. *Presupposition and implicature in compositional semantics*, pages 71–120.

Fox, D. (2018). Partition by exhaustification: Comments on dayal 1996.

Fox, D. (2020). Partition by exhaustification: Towards a solution to Gentile and Schwarz's puzzle. *Manuscript, MIT. https://semanticsarchive. net/Archive/TljZGNjZ.*

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition and Logical Form.* Academic Press.

Goldstein, S. (2019). Free choice and homogeneity. *Semantics and Pragmatics*, 12:23–EA.

Groenendijk, J. A. G. and Stokhof, M. J. B. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers.* PhD thesis, Univ. Amsterdam.

Hamblin, C. (1976). Questions in montague english. In Barbara H., P., editor, *Montague Grammar*, pages 247–259. Academic Press.

Humberstone, I. L. (1981). From worlds to possibilities. *Journal of Philosophical Logic*, pages 313–339.

Kamp, H. (1973). Free choice permission. *Proceedings of the Aristotelian Society*, 74:57–74.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and philosophy*, 1:3–44.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and philosophy*, 30:669–690.

Kratzer, A. and Shimoyama, J. (2002). Indeterminate pronouns: The view from japanese. In *3rd Tokyo Conference on Psycholinguistics*.

Križ, M. (2019). Homogeneity effects in natural language semantics. *Language and Linguistics Compass*, 13(11):e12350.

Marty, P., Romoli, J., Sudo, Y., and Breheny, R. (2021). Negative free choice. *Semantics and Pragmatics*, 14:13–EA.

Nygren, K. (2022). Free choice in modal inquisitive logic. *Journal of Philosophical Logic*, pages 1–45.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramotowska, S., Marty, P., Romoli, J., Sudo, Y., and Breheny, R. (2022). Diversity with universality. In Degano, M., Roberts, T., Sbardolini, G., and Schouwstra, M., editors, *Proceedings of the 23rd Amsterdam Colloquium*.

Roelofsen, F. and Farkas, D. F. (2015). Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, pages 359–414.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and philosophy*, 27:367–391.

Schwarz, F. (2013). Maximality and definite plurals-experimental evidence. In *Proceedings of Sinn und Bedeutung*, volume 17, pages 509–526.

Simons, M. (2005). Dividing things up: The semantics of or and the modal/or interaction. *Natural Language Semantics*, 13(3):271–316.

Tieu, L., Bill, C., and Romoli, J. (2019). Homogeneity or implicature: An experimental investigation of free choice. In *Semantics and linguistic theory*, volume 29, pages 706–726.

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., and Geurts, B. (2016). Scalar diversity. *Journal of semantics*, 33(1):137–175.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Willer, M. (2018). Simplifying with free choice. *Topoi*, 37(3):379–392.

Wright, G. V. (1968). An essay on deontic logic and the general theory of action.